

Analysing Irish prosody: A dual linguistic/quantitative approach

Ailbhe Ní Chasaide, Martha Dalton, Mika Ito, Christer Gobl

Phonetics and Speech Lab
Centre for Language and Communication Studies, Trinity College Dublin
anichsid@tcd.ie, daltonm@tcd.ie, mito@tcd.ie, cegobl@tcd.ie

Abstract

A project of Irish prosody is described which attempts to provide not only the basis for a linguistic description of the prosody of Irish dialects, but also the prerequisite quantitative characterization that is needed to allow us to use it for future technological applications, particularly text-to-speech development for Irish dialects. As with many other minority languages, there are particular challenges, but also particular opportunities to address. A multi-layered analytic approach is adopted, which will provide coverage of the three phonetic dimensions of prosody: pitch dynamics (intonation); voice quality; and temporal features. It is also envisaged that these analyses will provide the basis for an account that encompasses both the narrowly linguistic functions of prosody and its paralinguistic function of signaling attitude and emotion. In these last respects, this study aims also to contribute to the broader understanding of prosody, and to its modeling for more expressive speech synthesis. Given the relatively threatened status of Irish, we hope that by gearing our linguistic analysis to eventual technology exploitation, we can go beyond the mere documentation and aspire to the provision of tools that can support language teaching/learning and language usage generally.

Introduction

In this paper we describe a new project on the Prosody of Irish Dialects. Linguistic research on a minority language, in comparison to widely spoken and widely analyzed languages, often presents particular challenges and opportunities. As the approach adopted reflects many of these concerns it may thus be of general interest.

In our group we are keenly aware that knowledge concerning the linguistic structure of a language is a prerequisite for many speech technology applications. For example, in order to develop even a basic text-to-speech system, we ideally need models of the prosodic structure, as well as an understanding of segmental and grammatical structure etc. However, the analyses are often ill adapted to the eventual technological exploitation, and there is often a gap than can not easily be bridged. To be truly useful for technology, we need to adopt methodologies that ensure that the output can be harnessed by technology.

In the case of minority languages, speech technology is not just a luxury or gimmick. For endangered languages in particular, linguists are increasingly aware of the need to record and document them for posterity. Tools such as high quality text-to-speech facilities would serve to preserve models of the spoken language. Beyond the 'preserving' function, they could further play an important role in supporting language teaching/learning and language usage, crucial to the survival of the language.

The conundrum is that the non-commercial status of these languages makes it difficult for them to attract the funding and manpower for such linguistic research and technology adaptations. The lack of commercial viability arises from the limited size of the community of users, and/or as is the case with Irish, from the fact that most users are bilingual, so that product producers see the market as already served.

The need for Irish prosody research

Many of the goals of the present project have been shaped by these kinds of considerations, and it is felt that the linguistic analysis must maximize the potential downstream usefulness for later technology applications, particularly text-to-speech development. It therefore relates closely to another ongoing project, WISPR, which directly targets such technological developments (see parallel paper in this Workshop).

In setting out to describe Irish prosody, there are some further challenges which are shared with many other minority languages. Firstly, there is no prior account to draw on. Although there have been many accounts of the segmentals of Irish dialects (de Búrca, 1958; de Bhaldraithe, 1945), there is to date virtually no available coverage of segmental aspects, other than some short fragments on intonation (Blankenhorn, 1982; de Bhaldraithe, 1945). Consequently, the research is necessarily an exploration of mostly uncharted territory.

Another issue, common to many minority languages, is that given the historical and sociological context, there is no standard dialect. Rather, for Irish, we are confronted with four main dialects: and focusing on a single one involves making assumptions as to which one we regard as dominant. For that reason, in the present project, a cross-dialect approach has been adopted, and the analysis will be carried out on the four dialects of Donegal, Mayo, Connemara and Kerry Irish as illustrated in Figure 1.

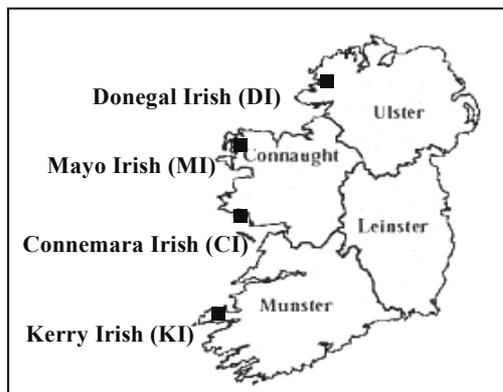


Figure 1: Map showing location of the four dialects of Irish, included in the analysis

One of the major linguistic interests of the descriptive material will be to show up what is common to all dialects and where the major divergences occur. The cross-dialect approach is dictated also partly by the fact that we are eventually interested in the provision of text-to-speech for all the main dialects.

It should be pointed out that research in minority languages also offers particular opportunities. First of all it allows us to reconsider many assumptions about language structure universals which are based on analyses of very few languages, mostly English. It is also an opportunity to look in a fresh way at theoretical issues and methodologies, and to try new perspectives, as one is in some sense “freed up” from current methodological straight jackets.

Project goals

As an approach to prosody description this project adopts a number of strategies, some of which are in line with current linguistic analyses, some of which are rather novel. The specific goals and methodologies are where relevant tailored to the situation of Irish and our hopes for technological downstream dividends. They also reflect some of the ongoing research preoccupations of our laboratory, and an interest in providing a more holistic account of prosody in general, which of course would have implications for the broader field of synthesis of any language. From this perspective, there are three main goals.

1. Integrating the three phonetic dimensions

It will describe the three phonetic dimensions of prosody: intonation, voice quality and rhythmic/temporal features. While most descriptive analyses of the prosody of individual languages provide an account of melodic structure, we feel that all three (in our view) phonetic correlates need to be incorporated where possible. Thus we will aim to provide an account of:

Intonation/melody: account of the primary pitch patterns of each dialect, and how these differ.

Voice quality: account of voice source correlates of major intonational categories, and an account of how voice quality is exploited for linguistic differentiation and for

paralinguistic communication of affect and attitude (see next subsection). This is largely absent from descriptions of prosody, but some of our own research (e.g. Ní Chasaide and Gobl 2004) as well as the work of other researchers in the field indicate that coverage of pitch dynamics without taking corresponding modulation of the voice source into account is inevitably a partial treatment of prosodic phenomena. This for us is very linked to some basic research on the voice which we are carrying out in the lab. It is also of course relevant to downstream synthesis applications, although this is in the longer term.

In technical terms, this is a difficult area to work in as obtaining reliable measurements is not at all straightforward. For this reason, voice quality analysis will be limited to two of the dialects.

Temporal/durational features: a description of the salient rhythmic/temporal characteristics of the dialects.

Although timing and rhythm is a major area of prosodic research, there is often a divide between it and that research focused on melody. To gain a proper understanding of how the prosody of a language works, we feel that a parallel description is required, so that we get some sense of how the temporal features contribute to those prosodic elements that are often described as involving pitch alone.

The need to include an account of temporal and rhythmic factors arises also from our long-term perspective of providing prosodic modeling for speech synthesis of Irish. Finally, within Irish linguistics, much of the interest will be focused on cross-dialect divergence and on the rather striking rhythmic differences in these dialects.

2. Encompassing linguistic/paralinguistic functions.

A second major goal is to provide an account that bridges the gap between the linguistic and paralinguistic functions of prosody. Early treatments of intonation such as O'Connor and Arnold (1961) saw the paralinguistic signaling of attitude and emotion as a primary role of prosody in speech communication, more recent generations of linguists have tended to shy away from this aspect, and focus more narrowly on its linguistic functions such as marking focus, phrase boundaries or even sentence mode (differentiation between declaratives and interrogatives).

While the initial thrust of our analyses will also target the linguistic level, we will be incorporating some analyses of affectively colored speech with a view to providing some initial model of how the linguistically relevant constituents of prosody vary with the attitude and emotion of the speaker.

This particular aspect of the project links to research we have been doing on the mapping of different voice qualities to affect (Gobl and Ní Chasaide, 2003) and on how voice quality and f_0 combine in affect signaling (Gobl et al 2002, Ní Chasaide and Gobl 2004). It is our firm view that the field of prosody has become falsely fragmented in a number of ways, and that we need to look at linguistic and paralinguistic phenomena within a single

framework. This is of course related to our intention of drawing together the different phonetic correlates of prosody: it is unlikely that one can make progress in describing the paralinguistic dimension without incorporating voice quality, along with pitch dynamics and temporal aspects.

Our research to date on voice quality, pitch and affect signaling was carried out through perception experimentation (references as above) with stimuli generated using formant based synthesis and a sophisticated voice source model (Fant et al., 1985). Our present project provide us with an opportunity to combine this with an analytic approach. This aspect of our research links to the objectives of a newly launched network of excellence on emotion (HUMAINE, 2004).

It also relates to possible downstream developments in speech synthesis, and in particular the provision of synthesized voices which are capable of conjuring specific emotions and attitudes (Gobl et al, 2002; Gobl and Ni Chasaide, 2002; Ni Chasaide and Gobl, 2002).

3. Qualitative and quantitative coverage

In keeping with the fact that we as linguists want to provide a linguistic account which is readily harnessed for technology applications, we are adopting a dual qualitative and quantitative approach. In the first, current phase, a linguistic, qualitative analysis of corpus materials is being carried out within the framework of Autosegmental-Metrical phonology (Ladd, 1996), using tone labels adapted from the ToBI annotation system. This analysis is an account of the possible combinations of tones (e.g. H*, L*+H) associated with accented syllables and phrase boundaries. The analysis is essentially auditory, and the researcher is guided by the visual display of f0. Figure 2 illustrates a ToBI-type analysis with f0 display for the sentence *Bionn ealaí ag snámh in Árainn Mhór*. This is the current standard for intonational analysis, and an account in this framework will be broadly accessible and allow comparison with other studies on languages and dialects elsewhere.

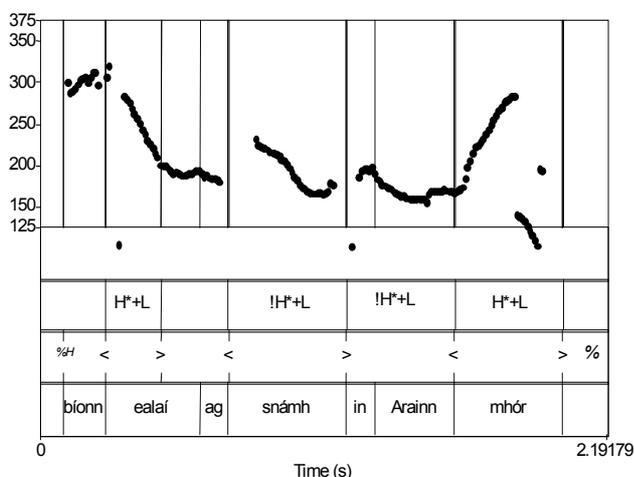


Figure 2: Display of Autosegmental-Metrical analysis of an Irish utterance, showing from top: f0 contour; tonal labels; prominence and boundary labels; and orthographic representation.

In the second phase of our intonation analysis, the same materials will be quantitatively analysed, using the Fujisaki model (Fujisaki, 1983). Figure 3 illustrates the same sentence, whose intonation is modeled in terms of Fujisaki parameters.

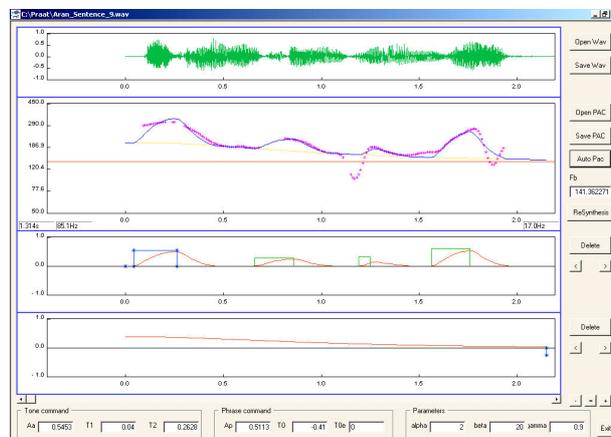


Figure 3: Display of Fujisaki analysis of the Irish utterance in Figure 2, showing from top: speech waveform; the original f0 contour (*dotted line); the modelled Fujisaki contour (solid line); the Fujisaki accent commands; and the Fujisaki phrase command.

The Fujisaki modeling is well adapted to synthesis applications and to our technology related goals for Irish. The quantitative measurements are also required by the second goal mentioned above, namely that of providing coverage of paralinguistic and linguistic phenomena within a single framework. As regards intonation, many of the prosodic manifestations of attitude and emotion concern not the configuration of tonal sequences, but rather changes in the dynamic range and the average level of f0 (Scherer, 1986). The Autosegmental-Metrical approach, using ToBI-type labels is highly abstract. While it is admirably suited to capture the linguistic, contrastive aspects of intonation, for which it was devised, it abstracts away much of the paralinguistic related variation.

Conclusion

In this project our goals and analytic strategies are motivated by different though not necessarily competing considerations:

- providing an account of an uncharted aspect of linguistic structure;
- providing an account of Irish intonation that is accessible to those interested in the typology of intonation systems;
- providing quantitative modeling of intonation and temporal structure that we can subsequently exploit in developing technology applications, particularly in the provision of text-to-speech for Irish
- providing basic research that will contribute towards a better understanding of the nature of prosody, and in particular of how the same set of phonetic dimensions (pitch, voice quality, and timing) are simultaneously exploited by the speaker to convey both linguistic and paralinguistic information.

To the extent that this last objective is achieved, this work will be contributing also to the development of more human-like expressive synthetic voices.

Acknowledgements

This research has been financially supported by a Government of Ireland Senior Research Fellowship to the first author, funded by the Irish Research Council for Research in the Humanities and Social Sciences, and by the research project *Prosody of Irish Dialects: the use of intonation, rhythm and voice quality for linguistic and paralinguistic signalling*, which is also funded by the Irish Research Council for Research in the Humanities and Social Sciences.

References

- Blankenhorn, V. S. (1982) Intonation in Connemara Irish: A Preliminary Study of Kinetic Glides. In *Studia Celtica* 16-17 (pp. 259-279).
- de Bhaldraithe, T. (1945) *The Irish of Cois Fhairrge*, Co. Galway, Dublin Institute for Advanced Studies.
- de Búrca, S. (1958) *The Irish of Tourmakeady*, Co. Mayo. Dublin Institute for Advanced Studies.
- Fant, G., Liljencrants, J. and Lin, Q. (1985). A four-parameter model of glottal flow. *STL-QPSR*, Speech, Music and Hearing, Royal Institute of Technology, Stockholm, 4, 1-13.
- Fujisaki, H (1983). Dynamic characteristics of voice fundamental frequency in speech and singing. In Peter F. MacNeilage (Ed), *The production of speech* (pp. 39–55). Springer, New York.
- Gobl, C. and Ní Chasaide, A. (2002). Dynamics of the Glottal Source Signal: Implications for Naturalness in Speech Synthesis. In E. Keller, G. Bailly, A. Monaghan, J. Terken and M. Huckvale (Eds.) *Improvements in Speech Synthesis* (pp. 273-283), Wiley and Sons.
- Gobl, C. and Ní Chasaide, A. (2003). The role of voice quality in communicating emotion, mood and attitude. *Speech Communication*, 40, 189-212.
- Gobl, C., Bennett, E. and Ní Chasaide, A. (2002). Expressive synthesis: how crucial is voice quality. In *Proceedings of the IEEE Workshop on Speech Synthesis*, Santa Monica, California, paper 52, 1-4.
- HUMAINE Network of Excellence (2004-2008). Funded by the EU Sixth Framework Programme. <http://emotion-research.net>
- Ladd, D. Robert (1996). *Intonational Phonology*. Cambridge University Press
- Ní Chasaide, A. and Gobl, C. (2002). Voice Quality and the Synthesis of Affect. In E. Keller, G. Bailly, A. Monaghan, J. Terken and M. Huckvale (Eds.) *Improvements in Speech Synthesis*, Wiley and Sons, 252-263.
- Ní Chasaide, A. and Gobl, C. (2004). Voice quality and f_0 in prosody: towards a holistic account. *Proceedings of the 2nd International Conference on Speech Prosody*, Nara, Japan.
- O'Connor, J.D. and Arnold, G.F. (1961). *The Intonation of Colloquial English*. Longman, London.
- Scherer, K.R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin*, 99, 143-165.